DOCUMENT RESUME

ED 119 487 FL 007 454

AUTHOR Lutz, Marjorie
TITLE The Development of Foreign Language Aptitude Tests: A Review of the Literature.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-TDR-66-4
PUB DATE Sep 67
NOTE 31p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage
DESCRIPTORS *Aptitude Tests; *Language Ability; Language Learning Levels; Language Proficiency; *Language Tests; *Literature Reviews; *Predictive Ability (Testing); Predictive Measurement; Second Language Learning; Test Construction; Test Reliability; Test Reviews
IDENTIFIERS *Language Aptitude; Modern Language Aptitude Test

ABSTRACT

Early language aptitude tests were generally tests of ability in English or work-sample tests in the target language or an artificial language. Recent work has involved various correlational studies and factor analyses to determine what factors correlate most with success in foreign language learning. Approaches include: (1) correlations of several language aptitude tests with the U.S. Air Force Schools English Comprehension Level Tests to study English learning ability in foreign students; (2) a predictor study using the vocabulary and paragraph reading sections of the Pennsylvania State College Academic Aptitude Examination, and (3) a study involving work-sample tests and artificial languages. Factor analytic studies seem of greater relevance now, to analyze and measure abilities constituting language aptitude. A study by Gardner and Lambert analyzed 24 variables of language skills and mental abilities and identified four as indices of second language skills. A study by Pimsleur, Stockwell and Comrey concluded that verbal intelligence and motivation were the main success factors. The Modern Language Aptitude Test consists of five subtests concerning number learning, phonetic script, spelling clues, words in sentences and paired associates. Reviews of the MLAT indicate that it is probably the most effective available, but criticize certain aspects. (CHK)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED119487

COLLEGE ENTRANCE EXAMINATION BOARD
RESEARCH AND DEVELOPMENT REPORTS

•

TEST DEVELOPMENT REPORT
TDR-66-4 SEPTEMBER 1967

The Development of Foreign Language Aptitude Tests: A Review of the Literature

•

Marjorie Lutz

Test Development Division, ETS

FL007454

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY
BERKELEY, CALIFORNIA

THE DEVELOPMENT OF FOREIGN LANGUAGE APTITUDE TESTS:

A REVIEW OF THE LITERATURE

Marjorie Lutz
Educational Testing Service
Test Development Division
Special Projects Department

Interest in foreign language aptitude testing originally resulted from the need for selection devices for both screening prospective students and assigning students to appropriate levels of study. This interest was high long before W. V. Kaulfers (1939) wrote that "No phase of foreign language research, with the possible exception of studies in the field of reading and achievement testing, has received attention from so many investigators as the measurement of linguistic aptitude or language talent." He went on to say that "foreign language achievement has been correlated with almost every ability and faculty conceivable to man" -- including, of course, intelligence. Although intelligence was once thought to be a principal contributor to success in learning a foreign language, it has been definitely shown that language aptitude consists of factors more specialized than general intelligence. J. B. Carroll (1962) wrote that "The proposition that foreign language aptitude is relatively specialized can be introduced by pointing out the well-known fact that intelligence tests have been largely unsuccessful in screening individuals for language training." He went on to say that, although certain cutting points may be introduced to eliminate those of limited intellectual ability, there are

apparently wide variations in the language-learning ability to those who are above the cutting point.

The article from which this was taken was included as a chapter in R. Glaser's Training Research and Education under the title "The Prediction of Success in Intensive Foreign Language Training" and presented some of the most recent and comprehensive work that has been done in the area of foreign language aptitude testing. Even though the purpose of this chapter was to report the major findings that are applicable to the screening of personnel for military and governmental programs for intensive and semi-intensive foreign language training, it contained much information that is relevant to the development of foreign language aptitude tests in general, including those which might be used for prediction in typical school courses.

According to Carroll the chapter is an attempt to demonstrate the truth of the two following propositions (a) that facility in learning to speak and to understand a foreign language is a fairly specialized talent (or group of talents), relatively independent of those traits ordinarily included under "intelligence" and (b) that a relatively small fraction of the general population seems to have enough of this talent to be worth subjecting to the rigorous, intensive, expensive training programs in foreign languages operated by military and governmental organizations, or by such private organizations as missionary societies, businesses, and industries engaged in overseas operations.

In order to put this entire problem in perspective, it is necessary to take a brief look at the historical development of foreign language aptitude tests. The material that follows is a summary of what is presented in Carroll's article. Reviews were presented in more detail by Henmon (1929) and by Carroll (1963).

Early efforts to develop aptitude tests for foreign languages resulted in tests generally of two sorts (a) tests of ability and achievement in the English language and (b) work-sample tests involving short lessons or problems in the language to be studied or in an artificial language. The tests that were developed presented essentially intellectual tasks to be solved by analytical procedures quite similar to those demanded by the kind of teaching that was prevalent in the 1920's and 1930's when the main objective of study was to teach the student to read or, indeed, just to translate a foreign language.[1]

These tests were unsatisfactory for a number of reasons: (a) They correlated highly with intelligence measures which could predict as well as the special prognostic tests. (b) They involved dependence on certain specific prior learnings such as knowledge of grammatical terminology and morphological processes like prefixing and suffixing. (c) They were incompatible with the changing teaching objectives which were moving away from primarily written knowledge of a language to the development of aural or oral abilities.

This brief look at the background of foreign language aptitude testing indicates the present need for accurate predictive instruments. And since the trend in research on language aptitude today seems to be toward some specialized factors that account for language-learning ability, most of the recent work in the field has involved various correlational studies and factor analyses to discover precisely what factors correlate most highly with success in foreign language learning.

Illustrative Studies

The studies that are reviewed in the following section present only a small sample of those that have been conducted and are included mainly to illustrate approaches that have been used to investigate foreign

language aptitude and aptitude testing. These approaches include (a) correlations of several language aptitude tests with the U. S. Air Force Schools English Comprehension Level tests to study English-learning ability in foreign students, (b) a prediction study using the vocabulary and paragraph reading sections of the Pennsylvania State College Academic Aptitude Examination, and (c) a study involving work-sample tests and artificial languages and another comparing an artificial language test with a specially constructed prognosis. In a later section, various factor analytic studies are reviewed in an attempt to delineate the nature of several specific factors common to linguistic aptitude and to examine the role of additional factors such as interest and motivation.

Prediction of English Language Achievement for Foreign Students

S. Sako and B. Fruchter (1965) investigated the utility in the prediction of the language-learning ability of foreign students of the English Comprehension Level (ECL) tests (which have been used at the U. S. Air Force Language School since 1954) in combination with the English language aptitude battery.

The English language aptitude battery of tests consisted of four predictor variables: (a) Sound Test, (b) Interest Test, (c) Word Study Test, and (d) Word Analysis Test. The Sound Test was designed to sample the ability of students to discriminate the English phonetic sounds. The Interest Test was developed to measure the interest and motivation of students relative to linguistic materials. The Word Study Test was constructed to sample the ability of students to memorize words by association with sounds and pictures. The Word Analysis Test was designed to determine the grammatical sensitivity of students by testing their ability to use words in sentences through inductive and deductive reasoning. (This reviewer was unable to obtain samples of

these tests.)

The entrance ECL scores were used also as the first criterion variable; fourth-week ECL scores were used as the second criterion variable, and the mean improvement units as the third criterion variable. The mean improvement unit was obtained by dividing the individual ECL improvement score by the mean ECL improvement score.

On the basis of the results, they concluded that when entrance ECL scores and the aptitude test scores were used as predictors, this combination seemed to give the maximum predictive utility for language preformance.

Vocabulary and Paragraph Reading Tests as Predictors. H. C. Peters (1953) reported a study concerned with the prediction of success and failure in elementary courses in French, Spanish, and German, using scores on the Pennsylvania State College Academic Aptitude Examination (Moore and Castore, 1948), parts one and two, Vocabulary and Paragraph Reading. Considering the importance of vocabulary in the learning of a foreign language, it was felt that the measurement of the skill involved in learning a vocabulary would be of some value in predicting success in learning a foreign language. Since this could not be done directly, a measure of proficiency in English vocabulary was used as being indicative of the ability to learn vocabulary. This assumes that the same skills are involved in learning the vocabulary of a foreign language.

The Paragraph Reading Test was used to indicate a subject's ability to understand the meaning of a paragraph. For this he "must have an adequate vocabulary, must be able to learn the meanings of new words from their contexts, and must have some knowledge of grammar."

The investigator sought a score on both parts of the Academic Aptitude Examination "below which are found those students who cannot

make passing grades ... in their language courses." The two parts of the test were used both separately and in combination in locating the different cut-off points for each language. The results showed that although there was differential predictive efficiency for the different languages there was very little difference in the efficiency of the predictive instruments themselves. The combined tests were generally most successful and the Vocabulary Test probably somewhat more effective than the Paragraph Reading Test.

Artificial Language Tests. Investigating, among other things, "To what extent we could predict success in the study of German by means of the Prognosis and Artificial Language tests, when a German, Yiddish, or Hunter College High School background was present." J. A. MacNaughton and M. R. Altenhein (1950) administered a specially constructed prognosis test to students entering the first semester of German study. The test was based on the Barry-Rice prognostic test in Spanish and included the following parts: (a) a test of the knowledge of English grammar, (b) application of rules of German grammar and word formation, and (c) translation of German sentences, based on analogy and the use of cognates and on vocabulary memorized from the previous page.

The Prognosis test was compared with an artificial language test (the third part of the American Council on Education Psychological Examination for College Freshmen -- 1935 and 1936 editions) in its ability to predict success as measured by several achievement tests and by instructors' grades.

The general conclusion that is relevant to the present report was that the Prognosis test appeared to be a better means of predicting success in German -- for the students studied.

The MacNaughton study was quite specialized both in its sample of students and in having developed its own prognosis test. A later study of artificial language tests, conducted by S. M. Sapon (1955), presents a more promising picture of the value of this form of work-sample testing.

Sapon first listed what he considered to be the primary desiderata in the design of a work-sample test: (a) It must reflect linguistic phenomena actually found in the languages of the world and must not be a completely analytic, English-like abstract code. (b) It must have a minimum of structural similarity to English. (c) The training and measurement techniques applied in the work-sample should approximate those found in actual course-work. Sapon wrote, "Herein should also be considered the role of various factors in language learning. Rote memory, while of obvious significance in some types of foreign language training, should not be depended upon to the exclusion of other factors. The growing acceptance of more direct methods of instruction call for an equivalent emphasis on inductive learning." (d) The "language" of the work-sample must, within the framework of previously stated needs, be of sufficiently compact structure to permit adequate instruction and practice in a short time period, yet still allow the constructor to manipulate varying degrees of complexity of the material. In addition, the problem of providing some degree of motivation towards optimum performance in the test must also be considered in the test design.

A complete description of Sapon's test, using the artificial language Temtem, was presented in the article. Results showed that total score on the test as the predictor variable correlated .53 with the criterion normalized academic grades. The reliability of the test was .68, as determined by Kuder-Richardson formula 20. At the time of the writing of the article, the test was being modified and the number

of items increased to raise its reliability. The author concluded that the results of the test's application to a student sample undergoing a trial course in Mandarin Chinese seemed to confirm the predictive value of this type of foreign language work-sample and to suggest its further development.

The studies above are cited as examples of some of the various ways investigators have gone about studying the prediction of success in foreign language courses. But to our present interest in the development of tests of foreign language aptitude, the factor analytic studies are of far greater relevance. H. T. Manuel (1960) wrote that although it is important to see how various tests fit together to discover differences in the ability to learn foreign languages, we should not be satisfied with prediction coefficients expressing only the relationship between a pretest and a measure of actual performance. "We need also to explore what students can do, what abilities they have for language learning." He advocated a shift from prediction to an analysis and measurement of the abilities that constitute aptitude for a foreign language. This is the aim of the studies now to be reviewed.

Factor Analytic Studies

In a study by R. C. Gardner and W. E. Lambert (1965), 24 variables were factor analyzed in an attempt to determine the factorial structure of a battery of tests designed to measure various language skills in order not only to clarify the relationship of intelligence to language aptitude and second-language achievement, but also to delineate the "specific second-language skills associated with specific language-learning abilities."

Variables 1-5 were the five MLAT subtests, Number Learning, Phonetic Script, Spelling Clues, Words in Sentences, and Paired Associates. Variables 6-10 were five of the tests from Thurstone and Thurstone's

Primary Mental Abilities test battery, including the Verbal Meaning, Space, Reasoning, Number, and Word Fluency tests. Variables 11-14 were subtests from the Cooperative French Listening Comprehension test, Form A which included Phonetic Discrimination, Answering Questions, Completion of Statements, and Completion of Passages. The Reading, Vocabulary, and Grammar subtests from the Cooperative French Test, Elementary Form Q constituted variables 15-17 and variables 18-21 were developed specially for the study: French Free Speech, French Reading Fluency, French Pronunciation Accuracy, and Standard French Accent. The remaining three variables were extracted from school records: Midterm French Grade, Final French Grade, and Academic Average, including all grades except French.

The 24-variable correlation matrix was factor analyzed using the Centroid method. Using the Varimax rotation solution, 7 orthogonal factors were extracted. To summarize the results, 4 factors were identified as indices of second-language skills, suggesting relatively independent dimensions of achievement. The first three were defined as "French Vocabulary Knowledge," "School French Achievement," and "Oral French Reading Skill," and the fourth tentatively labelled "Relative French Sophistication," the ability to perform with a minimum of translation, as though one understood the language.

A fifth factor, "Linguistic Reasoning," appeared to describe much of the variance common to the measures of language aptitude. It was present in all five MLAT subtests, and also in measures of language achievement and reasoning skills (Variables 8, 11, 15, and 17 -- Reasoning, Phonetic Discrimination, Reading, and Grammar).

Two additional factors, defined as "intelligence" and "Verbal Knowledge," were orthogonal to the other factors but were "of little interest" to the study. The report does not specify which variables had loadings on these factors.

The thesis of another study (Pimsleur, Stockwell, and Comrey, 1962) was that it is fallacious and even harmful to believe that a "special talent" is required to learn a foreign language. The authors wrote that their investigation "seeks to reduce the so-called 'talent for languages' to a set of well-defined, measurable components. Tests of the components will then be used to predict probable success in foreign language courses."

An important consideration of the study was the influences present in the typical high school and college courses that do not enter into intensive training courses and that need to be taken into account to attain high predictive validity. The article went on to say that prognostic tests composed entirely of intellectual tasks probably owed their rather low validity to their failure to include important nonintellectual variables, notably motivation and personality variables. In addition, in the light of the current shift toward teaching the spoken language, the study investigated the differential prediction of traditional (grammar-reading) and audio-lingual achievement.

Two separate studies for which the data consisted of scores on a number of predictor variables and several criteria, were conducted. The 21 predictor variables used in the first study were chosen to represent factors hypothesized to be related to success in learning a foreign language: associative memory, analytic reasoning, reasoning by analogy, physical dexterity in articulation, ability to change linguistic set, auditory discrimination, interest in foreign languages, and several others. The variables included parts I, III, IV, and V of the MLAT (Number Learning, Spelling Clues, Words in Sentences, and Paired Associates), a Letter Series test adapted from Guilford, Reading Aloud tests I and II (meaningful and meaningless material), Paraphrase, Rhymes, Synonyms, and Phonetic Discrimination tests, Linguistic Analysis tests adapted from Kabardian and a Guilford-

Zimmerman Verbal Comprehension test. Additional variables were age, sex, bilingualism, and high school language and math-science grades. The two criteria were French II final grades and a French Speaking Proficiency test.

A matrix of zero-order Pearsonian correlations among these variables was computed. From this, eight centroid factors were extracted and then rotated, using Kaiser's varimax criterion. The factors were described as follows: (a) Reasoning, which appeared to include the ability to induce the orderly principle from a group of examples; (b) Word Fluency, which included Paraphrase, Synonyms, and Rhymes; (c) Biographic (also including Paired Associates and Words in Sentences); (d) Achievement in French, made up of the two criteria; (e) Speed of Articulation, the Reading Aloud tests; (f) an unidentified bipolar factor including Letter Series, Phonetic Perception, and Linguistic Analysis I with negative loadings and Bilingualism with positive loadings (Bilinguals do poorly on the other three tests.); (g) Response Set, the tendency of some to avoid mistakes by working a little slower, while others work rapidly even at the risk of making mistakes, and (h) Verbal Knowledge.

A multiple correlation coefficient was computed for each of the two criteria and reduction was attempted in the size of the batteries. For predicting French grades, the minimum battery consisted of Number Learning (or Spelling Clues), Words in Sentences, Letter Series, Reading Aloud I, Paraphrase, Linguistic Analysis II, age, and high school math-science grades. This seven-test battery yielded a multiple correlation of .43. For predicting Speaking Test scores, the minimum battery consisted of Spelling Clues, Letter Series, Reading Aloud II, Verbal Comprehension, and Bilingualism. This five-test battery yielded a multiple correlation of .42.

In the second study, variables that had proved uninteresting were discarded and others were added to tap additional factors, including auditory factors, measured by a Chinese Pitch Perception test, the Seashore Pitch Test, and the Seashore Timbre Test, and an interest or motivation factor, measured by two Interest tests. The MLAT parts were dropped for copyright reasons, despite their good results in the first study. The three criteria were all new to this study. The Cooperative French Test was used as a measure of achievement in the reading and writing skills. Oral achievement was rated by a laboratory instructor who had listened to students and graded them in the laboratory once a week for a full semester. Achievement in understanding spoken French was measured by the Pictorial Auditory Comprehension Test, an objectively scored test.

From scores on the 22 variables, eight factors were again extracted and rotated. These were identified as (a) Achievement in French, (b) Speed of Articulation, (c) Reasoning, (d) Biographic, (e) Pitch Discrimination, (f) Word Fluency, (g) Interest in Languages, and (h) Timbre Discrimination.

The criterion of Cooperative French Test scores could be predicted to the extent of a multiple correlation of .673 when all 19 tests were used. The best six-test battery, including Verbal Comprehension, Chinese Pitch, both Interest tests, sex, and high school language grades, can be administered in a 50-minute class period; it yielded a multiple correlation of .652 (estimated shrinkage, $\underline{R}$ = .646). The two other criteria could also be predicted by a small battery of tests, but these multiple correlations were considerably lower, probably due to the lower reliability of the criteria.

The tests contributing to the prediction of Cooperative French Test scores represent factors of verbal intelligence, motivation, pitch dis-

crimination, and certain biographic elements. In their discussion, the authors reported that it had been hoped the prediction would be on the basis of intellectual factors, such as the ability to discriminate sounds, to induce grammatical principles, and so on. Instead, it appeared from these studies that the two biggest factors in foreign language achievement were the very general ones of verbal intelligence and motivation. They concluded; "This finding says that as far as language study in college is concerned, anyone will do well who is intelligent and wants to learn, regardless of such concerns as having 'a good memory,' and 'good reasoning powers.'"

It is interesting to note that in the prediction of the first criterion, the grammar-reading goal, three of the MLAT subtests were among the highest contributors: Words in Sentences (.047); Spelling Clues (.033); and Number Learning (.025). The Spelling Clues subtest was also the highest contributor to prediction of the oral achievement criterion, and accounted for .058 of the prediction. These facts will be relevant to a discussion of the MLAT which follows the next factor analytic study.

In conclusion, the factors involved were verbal intelligence and motivation, plus reasoning, pitch discrimination, and timbre discrimination. Substantial improvement of prediction, however, probably demands the inclusion of entirely new factors as predictors; among them, the personality of the student and the characteristics of the teacher are those which appear most promising and are most in need of research attention.

The conclusions reached by Pimsleur et al (1962) are further substantiated by Gardner and Lambert (1959), who write that "when measures of aptitude are correlated with grades in language courses, the validity coefficients show considerable variability from situation to situation even with tests developed through factor analytic methods, suggesting

that variables other than linguistic aptitude are involved." Gardner (1958) had previously written that "linguistic aptitude and social-motivation are factorially independent, suggesting that inclusion of both aptitude and social-motivational tests in a prognosis battery would greatly improve prediction of achievement." Gardner and Lambert (1959) go on to hypothesize that achievement in the learning of a second language is dependent upon the same type of motivation that was necessary in order to learn the native language, namely, the desire to become a member of a cultural group. They concluded that this type of motivation and "verbal intelligence" are two main factors in achievement.

In an article pursuing the conclusions and recommendations of the study cited previously (Pinsleur et al., 1962) Pinsleur reviews the experimental literature pertaining to "the factors within a student which may help or hinder him in his effort to learn a foreign language." These have been gathered under a number of headings, namely, (a) intelligence, (b) verbal ability, (c) pitch discrimination, (d) order of language study and bilingualism, (e) study habits, and (f) personality factors.

Pimsleur reviews these areas one by one and relates them to several of the studies previously described in the present review. The general conclusions that are relevant here are summarized below.

The areas which have been most completely investigated are intelligence and native language (English) verbal ability. Since these are known to be highly related, it may be well to think of them as a single factor of verbal intelligence. One of the few firm results of the research performed thus far is that this factor, verbal intelligence, appears to correlate about .45 with foreign language achievement. It is thus the largest contributing factor, but the correlation of .45 means that it accounts for only about 20 per cent of the variance. There still remains 80 per cent to be explained by other factors. A number of studies have dealt with the factor of motivation; they indicate a positive relationship

which may be as high as .40. If this is the case, then verbal intelligence and motivation together account for perhaps 35 per cent of the variance in foreign language achievement. Pimsleur concludes that the greater part of the variance in foreign language achievement remains to be investigated.

The Modern Language Aptitude Test

Leaving aside for the time being the importance of motivational, social, and personality factors to language learning ability, let us consider specific tests that have been developed and ways in which they might be improved. This review has shown that probably somewhere between five and eight "factors" have been described or identified as influencing language achievement. In spite of the many different variables used in the factor analytic studies described, there seems to be fairly good agreement among investigators as to what the factors are basically. In terms of developing a test to measure them, however, the problem is no simple one, for no matter how much agreement there may be regarding the basic abilities needed, there may be literally hundreds of ways of measuring the hundreds of different item types and procedures.

What is wanted, ideally, is a relatively objective, not-too-lengthy test which would measure all the relevant factors. The validity and accuracy of prediction of the Modern Language Aptitude Test developed by Carroll and Sapon and published by The Psychological Corporation in 1955.[3]

According to the 1959 edition of the MLAT Manual, the test provides an indication of an individual's probable degree of success in learning a foreign language; that is, speaking and understanding, and reading, writing, and translating the "modern" spoken languages and Latin and Greek. It was validated only for literates in Grade 9 and above who knew English with native or near-native fluency. The test consists of five relatively uncorrelated subtests which are described below.

Number Learning. By tape recording, the examinee is taught a simple artificial system of number expression utilizing nonsense syllables to represent the digits 0, 1, 2, 3, and 4. He is asked to write down the Arabic numerals for a list of two- and three-digit numbers in the artificial system, which are spoken at a fairly rapid pace on the tape. Factors measured by this subtest are memory and a special "auditory alertness."

Phonetic Script. The examinee learns a series of phonetic symbols for some of the phonemes of English by listening to the pronunciation of syllables recorded on tape while following the printed representations of the syllables in phonetic symbols on the test paper; after every five items the examinee goes back and is tested on the material just learned. After the 30-item learning period, there is a 30-item test in which the examinee must indicate, for each item, which of four phonemically-printed syllables is pronounced on the tape. All phonemes used on the test occur in English, and no fine phonetic discrimination is required. The Phonetic Script subtest measures sound-symbol association ability and memory for speech sounds.

Spelling Clues. This is an adaptation of the Turse Phonetic Association Test to objective scoring. The examinee chooses which of five words has the same meaning as the word represented in abbreviated form. Example: kataklzm = 1. mountain lion; 2. disaster; 3. sheep; 4. chemical reagent; 5. population. The test is highly speeded and is somewhat dependent on English vocabulary and related, to a lesser, with the Phonetic Script test. Carrol (1962) reported that "Ability to produce phonemes accurately and to mimic basic sentences seems most closely related to Spelling Clues."

Words in Sentences. Each item consists of a key sentence with a word or phrase printed in capital letters, followed by one or more sentences with

words and phrases underlined and numbered. The examinee is directed to pick the word or phrase in the second sentence that "does the same thing" in that sentence as the capitalized word does in the key sentence.

Example: He spoke VERY well of you.

Suddenly (1) the (2) music became quite (3) loud. (4)

This subtest has a close relationship with ability to understand grammar and to speak grammatically. It is not related to grammatical terminology.

Paired Associates. The examinee studies a list of 24 "Kurdish-English" vocabulary equivalents for two minutes, in the next two minutes he practices recalling the English meanings and in the final four minutes he completes from memory a multiple-choice test of the presented vocabulary. This measures the rote memory aspect of foreign language learning.

The Manual states that "The generally high validity of the MLAT seems clear evidence that it measures basic abilities essential to facile foreign language learning." In summary, these abilities are (a) rote memory, (b) auditory alertness, (c) the ability to form associations between orthographic symbols and speech sounds, (d) "verbal ability" or knowledge of the native English language, and (e) sensitivity to grammatical structure.

The factors that are included in the subtests of the MLAT fairly closely parallel the seven factors identified by Carroll in a factor analytic study (1958) with three exceptions. The most minor exception is a factor resembling French's Speed of Association factor, which involved only a small group of tests and did not prove valid in predicting success. The second is a factor that Carroll identifies as Linguistic Interest -- a specific motivation, interest, or facility with respect to linguistic materials -- the factor that was so important in the Pimsleur and the Gardner and Lambert studies. A third exception is a factor called Inductive Language Learning Ability. This was recognized by Carroll as being an important aspect of

foreign language aptitude although he wrote that "It is not measured to any appreciable degree by the tests in the final MLAT battery" (1962). This factor seems most likely to be measured by the work-sample or artificial language approach.

Several reviews of the MLAT indicate that it is probably the most effective foreign language aptitude test available today, but criticize certain aspects of the test and recommend various modifications. H. T. Manuel (1960) wrote that the reported reliability coefficients are generally good for both the total test and the short form -- of 20 odd-even coefficients in one table, 16 are between .90 and .94. The reliability coefficients of the component parts of the test are lower. He reported that the validity coefficients vary widely at every level, with the following approximate median values: .53 for both the total test and the short form in grades 9-11; .44 and .40 for the total test and short form in college; and .52 to .48 for the two forms in intensive language courses. He pointed out that "These coefficients, it will be recognized, are of the magnitude often found in correlating general scholastic aptitude with total college marks at the end of the first semester and are high enough to indicate that the test may be used effectively."

Of the opinion that aptitude must be viewed in terms of classroom implementation of foreign language instructional goals rather than in terms of language and language behavior per se, W. D. Fisher and B. B. Masia (1956 b) reported that predictive validities of the MLAT are generally .20 higher than those obtained for general ability and intelligence tests. "But," they added, "it is difficult to determine whether this overall increment in prediction is traceable to systematic psychological- and linguistic-oriented studies made by Carroll's Harvard Language Aptitude Project or to the more accidental effect of the test constructors' ingenuity in devising test models which are empirically more efficient for reasons which are not very clear." Their evaluation of the test suggested that it measures the student's

ability to recode English and gave the erroneous impression that language learning is in large measure a matter of the substitution of one set of words for another.[4] Their recommendations for improving the test called for the inclusion of non-English sounds in the taped portion and the use of non-English linguistic characteristics in all other parts of the test, plus some non-Roman symbolization.

M. F. Shaycroft's (1956) review discussed the use of part scores for predicting particular kinds of learning difficulties, mentioning that several of the parts are reliable enough for such use, but that interpretations would have to be cautious. J. W. Gallagher and R. E. Spencer (1964) compared the MLAT with the English Placement Examination and with the Academic Aptitude Examination (Moore and Castore, 1948) and reported that of the three tests used as predictors of final grades, only the MLAT yielded significant validity coefficients for some part scores. Total score correlations were not significant. Their conclusion was that MLAT was useful only if individual part scores or multiple correlations were used.

An article by W. F. Marwuardt (1961) was concerned with modifying the MLAT for use in selecting foreign students to study English. He acknowledged that the test as it stands can be given only to persons who know English well, but wrote that only parts III and IV (Spelling Clues and Words in Sentences) really require English. He suggested that pictures could be used in part V (Paired Associates) instead of English words. It would seem that the only way of saving Parts III and IV for use with foreign students would be to translate them into the native language of the examinee, paralleling the existing form.

A study actually attempting to modify the MLAT, in this case for blind students, was reported by Gardner (1965). The modifications were made usually without recourse to Braille; dot-answer sheets were used for responses and the test items were presented auditorily. In one study, the Cooperative Vocabulary Test, Form Q, was substituted for Spelling Clues and a Phonetic

Discrimination Test was substituted for Phonetic Script. Gardner hypothesized that, because of the type of instruction necessary with blind students, the increased reliance of the modified tests on memory skills would improve their validity. He obtained consistent predictive power with the Words in Sentences test and positive correlations for the Number Learning, Spelling Clues, and Vocabulary tests.

The MLAT, as it now stands, uses only five different tests, that is, five different types of questions. The results achieved by Gardner's modifications suggest the further investigation of various ways of measuring the factors associated with ability to learn a foreign language. The value of other kinds of tests or item types can be observed also in dome of the factor analytic studies.

A recent study, conducted by M. Lasch (1965) and concerned with the development of a prognostic test for foreign languages, use still other item types. The test was developed to aid in the selection of students for a five-year training course offered by the University of Chile for prospective teachers of high school English. The skills selected for testing were the ability to recognize and use certain "mechanisms" of language: aural discrimination of sounds; pitch and stress patterns other than native; the oral reproduction of the same, isolated or in patterns; the grasping of semantic extension other than native; and the grasping and manipulating of structural devices such word order, word form, and function words. Lasch's paper is a description of how he tried to test these abilities. In all, the test consisted of eight different sections. Two of the more interesting and different item types are described briefly below. All stimulus material was recorded.

One type of item presented the student with a series of facts in 24 words consisting of a subject, a relative clause, and a predicate in the passive voice, and including a surname that is not too common and two

expressions of time (in the student's native language, Spanish). The rationale for this item type was as follows: "It was thought that any student, particularly one of language, should be able to decode, store and reproduce a reasonable amount of verbal information. This hearing and reproducing of long utterances is frequently indispensable in aural-oral drilling. The vocabulary used was the kind that any college student was likely to hear, yet not too colloquial."

Another type of item is illustrated by The Vernacular Mispronounced. "Each version consists of two utterances; in the first, the quality of three consonants is altered; in the second, the quality of four vowels is altered, in such a way as to produce consonants and vowels typical of English, in an utterance in Spanish. It is the kind of Spanish that might be heard from an untrained native speaker of English. In the first place, a certain amount of acoustic discrimination on the part of the student is required; secondly, he must be able to mimic the sounds as best he can."

Because of the small number of students in the sample, and other factors such as financial difficulties influencing a student's progress, no correlations have been made with success or failure. However, after one year, these general results were observed: Although the students who obtained the highest scores on the aptitude test did not necessarily obtain the highest marks in their regular course, the candidates with the lowest scores on the aptitude test failed the course. The section of the test that seemed to correlate best with the results obtained after the first year of instruction was hearing and reproducing nonsense words. These items consist of three consonant clusters all grouped in sequences in a way that does not occur in Spanish. (Example: /tu:l'bi:/ /'eknarf/ /swi:'xa:la/)

The present MLAT subtests seem to be the best compact and objective measures of foreign language aptitude available at this time. It would seem that the most appropriate direction in which testing in this area

should proceed -- ignoring, for present purposes, multiple prediction coefficients based on various aptitude tests, intelligence measures, and grades in other school courses -- is toward the development of a test similar to the MLAT but, hopefully, still more accurate in its predictive power and perhaps more broadly based in its content sampling and more varied in its ways of tapping language abilities.

Further investigation will have to be made regarding such a test's reliance on the English language background of the examinee. The MLAT is fairly consistent for predicting success in different languages. If the ideal of a "culture-free" test should prove at all feasible, it would be most useful to have a foreign language aptitude test that could be used with people of all linguistic backgrounds. In this case, all efforts should be directed toward freeing a new test from any reliance on English. If, on the other hand, such an ideal is either impossible or not suitable to the specific purposes of the test developer, there would seem to be little harm in the presence of material oriented to the English language. This would not be in disagreement with the viewpoint of those who feel that learning a second language is closely associated with native language ability.

A test which could more accurately report part scores for specific factors would be immensely helpful in predicting specific learning difficulties a student might encounter or in diagnosing certain difficulties for students already enrolled in language courses. The artificial language approach seems to merit further investigation, especially in view of the progress now being made with programmed instruction and computer teaching aids.

The preceding suggestions concerning the development of a foreign language aptitude test have been made with the assumption that any new test would be based on existing knowledge and on existing tests, the goal being to achieve improved accuracy of prediction with basically the same materials that have been used. Any different approaches would require intensive research to

further explore the abilities used in learning a second language and as yet unknown ways of measuring them. Studies would have to be carried out that were even more extensive than those of Carroll or Pimsleur. Therefore, if the aim of the test developer were simply to attempt to develop an additional test with more predictive accuracy than the MLAT, the task seems feasible and would primarily involve the examination of existing item types or new ones based on them with a view toward maximizing the predictive power of the test.

FOOTNOTE

[1] Tests developed during this period: (1) the Iowa Foreign Language Aptitude Examination (Stoddard & Vanderbeke, 1925), (2) the Luria-Orleans Modern Language Prognosis Test (Luria & Orleans, 1928), (3) the George Washington University Language Aptitude Test (Hunt, Wallace, Doran, Buynitzky, & Schwartz, 1929), and (4) the Symonds Foreign Language Prognosis Test (Symonds, 1930).

[2] For additional research on this subject see Richards & Appel (1956).

[3] Most of the literature pertaining to background studies for the development of the MLAT is cited under J. B. Carroll in the reference section of this report.

[4] On the subject, Dunkel (1948) wrote of the importance of the "transfer value" of the native language and the role of analogy in learning a second language, concluding that "Until [the learner] knows, from correction, the <u>full</u> range of what <u>cannot</u> be substituted or combined in the foreign language, he continues to base his substitutions and combinations on parallels he knows in this native tongue."

Modern Language Aptitude Test

Agard, F. B. & Dunkel, H. B. An investigation of second-language teaching. Boston: Ginn & Co., 1948.

Asher, J. J. Vision and audition in language learning. Percept. mot. Skills, 1964, 19, 255-300 (Monograph Supplement 1-V19).

Birkmaier, Emma M. An investigation of the outcomes in the eclectic reading and the modified Army method courses in the teaching of a second language. Unpublished doctoral dissertation, Univer. of Minnesota, 1949.

Birkmaier, Emma M. Modern languages. In C. W. Harris (Ed.), Encyclopedia of educational research. (3rd ed.) New York: Macmillan, 1960. Pp. 861-888.

Brooks, N., et al. Definition of language competences through testing. In F. D. Eddy (Ed.), Reports of Working Committees, 1959 Northeast Conference on Teaching Foreign Languages. Washington, D. C.: Georgetown Univer., 1959, Pp. 49-56.

Carroll, J. B. The Harvard foreign language aptitude tests. Yearb. Nat. Council Measmt Used Ed., 1955, 12 (2), 9-11.

Carroll, J. B. A factor analysis of two foreign language aptitude batteries. J. gen. Psychol., 1958, 59, 3-19.

Carroll, J. B. The prediction of success in intensive foreign language training. Harvard University Graduate School of Education, Laboratory for Research in Instruction, 1960. (Mimeo.)

Carroll, J. B. The prediction success in intensive language training. In R. Glaser (Ed.), Training research and education. Pittsburgh: Univer. of Pittsburgh Press, 1962. Pp. 87-136.

Carroll, J. B. Research on teaching foreign languages. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally & Co., 1963. Pp. 1060-1100.

Carroll, J. B., Sapon, S. M., and Richards, S. E. Construction and validation of a test battery for predicting success in spoken language courses. The Harvard Language Aptitude Project, Report No. 1, June 1954. (Hectographed for private circulation.)

Carroll, J. B. & Sapon, S. M. Prediction of success in a work-sample course in Mandarin Chinese. Amer. Psychologist, 1955, 10, 492-493.

Carroll, J. B. & Sapon, S. M. Psi Lambda foreign language aptitude battery. Cambridge: John B. Carroll, 1955.

Carroll, J. B. & Sapon S. M. Manual, Modern language aptitude test. (1959 ed.) New York: Psychological Corp., 1959.

Cieutat, V. J. Predicting speed of serial and paired-associate learning. Psychol. Rep., (December) 1963, 13, 786.

Dorcus, R. M., Mount, G. E., and Jones, Margaret H. Construction and validation of foreign language aptitude tests. Los Angeles: Department of Psychology, Univer. of California, 1952.

Doyle, H. G. Modern foreign languages: A chronicle of achievement. Mod. Lang. J., (October) 1956, 20, 269-296.

Ferster, C. B. and Sapon, S. M. An application of recent developments in psychology to the teaching of German. Harvard educ. Rev., 1958, 26, 58-69.

Fisher, W. D. and Masia, B. B. Foreign language prognosis tests [355]. In O. K. Buros (Ed.), The Sixth Mental Measurements Yearbook. Highland Park, N. J.: Gryphon Press, 1956a. Pp. 632-633.

Fisher, W. D. and Masia, B. B. Modern language aptitude test [257]. In O. K. Buros (Ed.), The Sixth Mental Measurements Yearbook. Highland Park, N. J.: Gryphon Press, 1956b. Pp. 632-633.

Firth, J. R. Selection for language training by a trial course. Georgetown Univer. Monogr. Ser. Lang. Linguistics, 1953, 4, 10-15.

Gallagher, J. W. and Spencer, R. E. Prediction of success in basic German at the college level. Ed. psychol. Measmt, (Winter) 1964, 24, 955-960.

Gardner R. C. Social factors in second-language acquisition. Unpublished thesis, McGill University, 1958.

Gardner R. C. Motivational variables in second-language acquisition. Unpublished doctoral dissertation, McGill University, 1960.

Gardner, R. C. and Lambert, W. E. Motivational variables in second-language acquisition. Canad. J. Psychol., 1959, 13, 266-272

Gardner, R. C. and Lambert, W. E. Language aptitude, intelligence, and second-language achievement. J. ed. Psychol., (August) 1965, 56, 191-199.

Guilford, J. P. and Christensen, P. R. A factor-analytic study of verbal fluency. Univer. sth. Calif. psychol. Lab. Rep., 1956, No. 17.

Harding, F. D. Language aptitude tests as predictors of success in a trial Mandarin Chinese course. Technical Memorandum PL-TM-56-8, Personnel Laboratory, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas, 1956a.

Harding, F. D. Language aptitude tests as predictors of success in a trial Russian course. Technical Memorandum PL-TM-56-5, Personnel Laboratory, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas, 1956b.

Harding, F. D. Tests as selectors of language students. Mod. Lang. J., (March) 1958, 42, 120-122.

Hascall, E. O. Predicting success in high school foreign language study. Personnel Guid. J., (December) 1961, 40, 361-367.

Henmon, V. A. C. Recent developments in the construction, evaluation, and use of tests in the modern foreign languages. In A. Coleman (Ed.), Experiments and Studies in Modern Language Teaching. Chicago: Univer. of Chicago Press, 1934. Pp. 191-218.

Henmon, V. A. C., Bohan, J. E., and Brigham, C. C. Prognosis tests in the modern foreign languages. New York: Macmillan, 1929.

Iowa Placement Examinations. Extension Division, University of Iowa, Iowa City, Iowa.

Kaulfers, W. V. Present status of prognosis in foreign language. Sch. Rev., 1931, 39, 585-596.

Kaulfers, W. V. Prognosis and its alternatives in relation to the guidance of students. Germ. Quart., 1939, 12, 81-84.

Kaulfers, W. V. Foreign language prognosis test [1940]. In O. K. Buros (Ed.), The 1940 Mental Measurements Yearbook. Highland Park, N. J.: The Mental Measurements Yearbook, 1941. Pp. 157-158.

Kjeldergaard, P. M. Predicting paired-associate learning speed. Psychol. Rep., (October) 1962, 11, 353-354.

Kurlander, E. D. The modern language aptitude test as a guidance tool. Sch. Counselor, (March) 1963, 10, 129-131.

Lado, R. Linguistics across cultures: Applied linguistics for language teachers. Ann Arbor: Univer. of Michigan Press, 1957.

Lado, R. Language testing: The construction and use of foreign language tests. London: Longmans, 1961.

Lambert, W. E. Measurement of the linguistic dominance of bilinguals. J. abnormal. soc. Psychol., 1955, 50, 197-200.

Lambert, W. E. Developmental aspects of second-language acquisition. J. soc. Psychol., 1956, 43, 83-104.

Lambert, W. E., Gardner, R. C., Olton, R., and Tunstall, K. A study of the roles of attitudes and motivation in second-language learning. Montreal: McGill Univer., 1961.

Lasch, M. A prognostic test of language learning ability. Unpublished study, Regional College in Temuco, Universidad de Chile, December 1965.

Luria, M. A. and Orleans, J. S. Luria-Orleans Modern Language Prognosis Test. Yonkers: World Book Co., 1928, 1930.

MacNaughton, Jacqueline A. and Altenhein, Margarete R. Investigation of prognosis in German. Mod. Lang. J., 1950, 34, 553-560.

Manuel, H. T. Modern language aptitude test. Personnel and Guid. J., (March) 1960, 38, 582-584.

Marquardt, W. F. Can foreign student selection be based on aptitude for learning English? Inst. Int. Ed. N. Bull., (April) 1961, 36, 2-8.

Michael, W. B. Foreign language prognosis test [232]. In O. K. Buros (Ed.), The Fourth Mental Measurements Yearbook. Highland Park, N. J.: Gryphon Press, 1953. Pp. 348-349.

Miller, G. A. Free recall of redundant strings of letters. J. exp. Psychol., 1958, 56, 485-491.

Moore, B. V. and Castore, G. F. The Pennsylvania State College Academic Aptitude Examination, 1947 revision. The Pennsylvania State College, 1948.

Mueller, K. A. and Wiersma, W., Jr. Correlation of foreign language speaking competency and grades in ten midwestern liberal arts colleges. Mod. Lang. J., (December) 1963, 47, 353-355.

Noble, C. E. and McNeely, D. A. The role of meaningfulness (m) in paired-associate verbal learning. J. exp. Psychol., 1957, 53, 16-22.

Peters, H. C. Prediction of success and failure in elementary foreign language courses. J. ap. Psychol., (June) 1953, 37, 178-181.

Pimsleur, P., Stockwell, R. P., and Comrey, A. L. Foreign language learning ability. J. ed. Psychol., (February) 1962, 53, 15-26.

Pimsleur, P. Experimental design in the language field. Mod. lang. Forum, 1957, 42, 157-163.

Pimsleur, P., Mosberg, L., and Morrison, A. L. Student factors in foreign language learning. Mod. Lang. J. (April) 1962, 46, 160-170.

Richards, S. E. and Appel, Joan E. The effects of written words in beginning Spanish. Mod. Lang. J. 1956, 40, 129-133.

Sako, S. and Fruchter, B. Statistical study of language achievement of foreign students. Lang. Learn., 1965, 15 (1/2), 43-49.

Salomon, Ellen Generation of prognosis testing. Mod. Lang. J., (October) 1954, 38, 299-303.

Sapon, S. M. A work sample test for foreign language prognosis. J. Psychol., 1955, 39, 97-104.

Seagoe, May V. Prediction of achievement in foreign languages. J. appl. Psychol., 1938, 22, 632-640.

Shaycroft, Marion F. Modern language aptitude test [357]. In O. K. Buros (Ed.) The Sixth Mental Measurements Yearbook. Highland Park, N. J.: Gryphon Press, 1956. Pp. 635-636.

Siciliano, E. A. An experiment in listening comprehension. Mod. Lang. J., 1959, 43, 226-227.

Starr, W. H. Competency first: New tests in foreign languages. In J. B. Carroll (Ed.) Proc. 1960 Invit. Conf. on Testing Problems. Princeton, N. J.: Educational Testing Service, 1961. Pp. 97-110.

Symonds, P. M. A foreign language prognosis test. Teachers College Record, XXI, March 1930.

Tharp, J. B. How shall we section beginning foreign language classes? Mod. Lang. J., 1929, 13, 433-449.

Thurstone, L. L. and Thurstone, Thelma G. Manual of the Chicago tests of primary mental abilities, ages 11 to 17. Washington, D. C.: American Council on Education, 1941.

Veon, Dorothy H. The relationship of learning factors found in certain modern foreign-language aptitude tests to the prediction of shorthand achievement in college. Stillwater, Okla.: Division of Commerce, Oklahoma Agricultural and Mechanical College, 1950. Pp. 74.

von Wittich, Barbara. Prediction of success in foreign language study. Mod. Lang. J., (May) 1962, 46, 208-212.

Weaver, W. W. and Kingston, A. J. A factor analysis of the Cloze procedure and other measures of reading and language ability. J. Commun., (December) 1963, 13, 252-261.